



UNIVERSIDAD  
DE GRANADA

Escuela Técnica Superior de Ingenierías Informática y de  
Telecomunicación  
Facultad de Ciencias

DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y  
MATEMÁTICAS

TRABAJO DE FIN DE GRADO

# Optimización de redes neuronales

Presentado por:

Blanca Cano Camarero

Tutores:

Juan Julián Merelo Guervós

*Arquitectura y tecnología de computadores*

Francisco Javier Meri de la Maza

*Análisis matemático*

Curso académico 2021-2022



# Optimización de redes neuronales

Blanca Cano Camarero

Blanca Cano Camarero *Optimización de redes neuronales.*

Trabajo de fin de Grado. Curso académico 2021-2022.

**Responsables de  
tutorización**

Juan Julián Merelo Guervós  
*Arquitectura y tecnología de computadores*

Francisco Javier Meri de la Maza  
*Análisis matemático*

Doble Grado en Ingeniería  
Informática y Matemáticas

Escuela Técnica Superior  
de Ingenierías Informática  
y de Telecomunicación  
Facultad de Ciencias

Universidad de Granada

## Resumen

**Objetivo informática:** Elegir un framework común para trabajar con redes neuronales así como una serie de problemas de complejidad media, tales como spambase. Establecer una línea base examinando los resultados obtenidos con la configuración base a la hora de entrenar este tipo de redes neuronales y el resultado obtenido. A partir de esa línea base, testear las diferentes restricciones, cambios en representación y suposiciones deducidos en la parte matemática para ver qué influencia tienen en la velocidad, en el resultado, o en ambos.

**Objetivo matemáticas:** El objetivo de esta parte es doble, en primer lugar, se propone analizar con detalle las demostraciones de algunos resultados de aproximación universal de redes neuronales para funciones continuas. En segundo lugar se propone realizar un estudio de la posible optimización de redes neuronales concretas en base a los resultados obtenidos empíricamente en la parte informática. Se tratará de modelizar matemáticamente dichos resultados y de obtener mejoras en la convergencia de las aproximaciones imponiendo, si es necesario, hipótesis más restrictivas en algunos de los elementos de las redes neuronales que se correspondan con su uso en la práctica.

**Libros:** [1] Abu-Mostafa, Y.S. et al.: Learning From Data. AMLBook, 2012. [2] G. Cybenko, Approximations by superpositions of a sigmoidal function, Math. Control Signal Systems 2 (1989), 303-314. [3] J. Conway, A Course in Functional Analysis, 2nd Edition, Springer-Verlag, 1990. [4] A. Géron, Hands-on machine learning with Scikit-Learn, Keras and TensorFlow: concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly, 2019. [5] K. Hornik, M Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989), 359-366. [6] W. Rudin, Real and complex analysis. McGraw-Hill Book Co., New York-Toronto, Ont.-London 1966.

**PALABRAS CLAVE:** redes neuronales LSTM series temporales selección de modelos validación selección de hiperparámetros detección de anomalías detector perturbación



## Summary

**KEYWORDS:** neural networks LSTM time series model selection validation hyper-parameters selection anomaly detection detector perturbation

## Índice general

<b>1</b>	<b>Introducción</b>	<b>11</b>
<b>I</b>	<b>Matemáticas</b>	<b>13</b>
<b>2</b>	<b>Polinomios de Bernstein</b>	<b>15</b>

## Índice de figuras

## Índice de tablas







## 1. Introducción

**Definición 1.1** ( Polinomios de Bernstein). Dada cierta función  $f : [0, 1] \rightarrow \mathbb{R}$ , se define el  $n$ -ésimo polinomio de Bernstein para  $f$  como

$$B_b(x) = B_n(x; f) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$



# **Parte I.**

## **Matemáticas**

Esto es un texto nuevo



## 2. Polinomios de Bernstein

**Definición 2.1** (Polinomios de Bernstein). Dada cierta función  $f : [0, 1] \rightarrow \mathbb{R}$ , se define el  $n$ -ésimo polinomio de Bernstein para  $f$  como

$$B_n(x) = B_n(x; f) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

La intuición que se esconde tras esta definición es la siguiente: Se pretende aproximar la función  $f$  a través de los puntos  $\frac{k}{n}$  con  $n \in \mathbb{N}$  fijo y  $k \in \{0, \dots, n\}$ . De tal forma que si evaluamos  $B_n(x)$  con  $x$  lo suficientemente próximo a  $\frac{k}{n}$  la imagen de  $B_n(x)$  se acerque a  $f\left(\frac{k}{n}\right)$ .

Para ello recordaremos el teorema del Binomio de Newton:

**Teorema 2.1** (Binomio de Newton). *Cualquier potencia de un binomio  $x + y$  con  $x, y \in \mathbb{R}$ , puede ser expandido en una suma de la forma*

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$$

Así pues en virtud de esta igualdad y puesto que nuestro dominio de definición de  $f$  es  $[0, 1]$ , para cualquier  $x \in [0, 1]$ .

Tenemos

$$1 = (x + (1 - x))^n = \sum_{k=0}^n \binom{n}{k} x^k (1 - x)^{n-k}$$

Multiplicamos ahora en ambos lados por  $f(x)$

$$f(x) = \sum_{k=0}^n f(x) \binom{n}{k} x^k (1 - x)^{n-k}$$

Y tenemos que la diferencia entre  $f(x)$  y  $B_n(x)$  es

$$f(x) - B_n(x) = \sum_{k=0}^n \left( f(x) - f\left(\frac{k}{n}\right) \right) \binom{n}{k} x^k (1 - x)^{n-k}$$

Que en valores absolutos resulta

$$|f(x) - B_n(x)| = \sum_{k=0}^n \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1 - x)^{n-k}$$

Observando esta ecuación 2 se desprende como es natural un teorema de convergencia.

**Teorema 2.2** (Teorema de aproximación de Bernstein). *Sea  $f$  una función continua en un intervalo  $I$  con imágenes en los reales. La secuencia de polinomio de Bernstein 2.1 converge uniformemente a  $f$  en  $I$ .*

## 2. Polinomios de Bernstein

Recordaremos antes la definición de convergencia uniforme:

**Definición 2.2** (Convergencia uniforme para funciones reales). Dado  $E$  un conjunto y  $\{f_n\}_{n \in \mathbb{N}}$  una sucesión de funciones de  $E$  a los reales; se dice que dicha sucesión converge uniformemente si para cualquier  $\varepsilon > 0$  existe un número natural  $m$  tal que para todo  $x \in E$  y cualquier natural  $n$  que cumpla  $n \geq m$  se tiene que

$$|f_n(x) - f(x)| < \varepsilon$$

Comencemos pues con la demostración 2.2

*Demostración.* Sin pérdida de generalidad supondremos que  $I = [0, 1]$ , como veremos esto no es restrictivo ya que si  $I$  fuera un intervalo cerrado existiría un homeomorfismo  $H$  tal que  $H^*(I) = [0, 1]$  y podríamos trabajar con  $H \circ f$  la cual respetaría todos los argumentos utilizados en la demostración.

Si  $I$  fuera un abierto consideraríamos su cierre y aplicaríamos el razonamiento anterior. De esta manera los supremos e ínfimos se mantendrían, ahora como mínimos y máximos y no se alteraría de ninguna manera la continuidad.

Tras la aclaración anterior podemos comenzar.

Sea  $\varepsilon > 0$  queremos probar que existirá un  $m_\varepsilon \in \mathbb{N}$  tal que para cualquier  $x \in I$  e  $n \geq m_\varepsilon$  se tenga que  $|f(x) - B_n(x)| < \varepsilon$ .

Para ello por estar  $f$  definida en un intervalo, se tienen dos consecuencias claves:

1. Está acotada, supongamos por  $M \in \mathbb{R}$ , esto es  $|f(x)| \leq M$ .
2. Es uniformemente continua, es decir: para cualquier  $\varepsilon > 0$  existirá un  $\delta_\varepsilon$  tal que para cualesquiera  $x, y \in I$  que cumplan  $|x - y| < \delta_\varepsilon$  entonces  $|f(x) - f(y)| < \varepsilon$

Dado  $N \in \mathbb{N}$  fijo pero arbitrario, para  $\{k | k \in \{1, \dots, N\} \frac{k}{N} \in I$  y tomando  $x \in I$  podemos acotar

$$|f(x) - f\left(\frac{k}{N}\right)| \leq |f(x)| + |f\left(\frac{k}{N}\right)| \leq 2M$$

Fijado un  $x$  del dominio, se tienen las siguientes particiones de índices:

$$\mathcal{A}_{x,N} = \{k | k \in \{1, \dots, N\} \text{ y } |x - \frac{k}{N}| < \delta_\varepsilon\}$$

Donde el  $\delta_\varepsilon$  se ha obtenido de la observación 2 tomando como  $\varepsilon$  el buscado para la convergencia uniforme.

Sea  $\varepsilon > 0$  y tomemos  $\delta_\varepsilon$  el delta de la definición de uniformemente continuo.

Por otro lado se define

$$\mathcal{B}_{x,N} = \{1, \dots, N\} - \mathcal{A}_{x,N}$$

Podemos elegir  $m$  de manera conveniente de tal forma que

$$n \geq \sup \left\{ (\delta_\varepsilon)^{-4}, \frac{M^2}{\varepsilon^2} \right\}$$

y agruparemos los términos de la sumatoria en dos partes, para los que  $|x - \frac{k}{n}| < n^{-\frac{1}{4}} \leq \delta_\varepsilon$ . Para éstos, se tiene por Newton



$$\sum_{k=1}^n \varepsilon \binom{n}{k} x^k (1-x)^{n-k} \leq \varepsilon \sum_{k=1}^n \binom{n}{k} x^k (1-x)^{n-k} = \varepsilon$$

Para el resto de términos, aquellos cuya  $k$  cumpla que  $|x - \frac{x}{n}| \geq n^{-\frac{1}{4}}$ , se tiene que  $|x - \frac{x}{n}| \geq n^{-\frac{1}{2}}$ .

Queda el resto de subíndices resulta acotada por, deberemos de ver que para una determinada  $N$ , esa desigualdad se queda en  $\varepsilon$

$$\sum_{k \in \mathcal{B}_{x,N}} 2M \binom{n}{k} x^k (1-x)^{n-k}$$

Para acotar esto nos va a costar trabajar un poco más

$$\begin{aligned} & \sum_{k \in \mathcal{B}_{x,N}} 2M \binom{N}{k} x^k (1-x)^{N-k} \\ &= 2M \sum_{k \in \mathcal{B}_{x,N}} \frac{(x - \frac{k}{N})^2}{(x - \frac{k}{N})^2} \binom{N}{k} x^k (1-x)^{N-k} \\ &\leq 2M \sqrt{N} \sum_{k \in \mathcal{B}_{x,N}} (x - \frac{k}{N})^2 \binom{n}{k} x^k (1-x)^{N-k} \end{aligned}$$

Tengamos ahora presenten las siguientes igualdades

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} = \frac{k}{n} \binom{n}{k} x$$

Partiendo de 2

□



## **Agradecimientos**

Agradezco a